2014-07-03

Best binary equivocation code construction for syndrome coding

Zhang, K

10.1049/iet-com.2013.0889
IET Communications
Institution of Engineering and Technology (IET)

Best binary equivocation code construction for syndrome coding

K. Zhang

Instituto de Telecomunicações

Dep. de Ciência de Computadores

Universidade do Porto

{kezhang}@dcc.fc.up.pt

M. Tomlinson, M. Z. Ahmed, M. Ambroze

University of Plymouth

United Kingdom

{M.Tomlinson, M.Ahmed, M.Ambroze}@plymouth.ac.uk

M. R. D. Rodrigues

Dep. of Electronic and Electrical Engineering

University College London

United Kingdom

{m.rodrigues}@ucl.ac.uk

Abstract

Traditionally, codes are designed for an error correcting system to combat noisy transmission channels and achieve reliable communication. These codes can be used in syndrome coding, but it is shown in this paper that the best performance is achieved with codes specifically designed for syndrome coding. In the view of the security of the communication, the best codes are the codes, which have the highest value of an information secrecy metric, the equivocation rate, for a given code length and code rate and are well packed codes. A code design technique is described, which produces the best binary linear codes for the syndrome coding scheme. An efficient recursive method to determine the

equivocation rate for the Binary Symmetric Channel and any linear binary code is also presented. A large online database of best equivocation codes for the syndrome coding scheme has been produced using the code design technique with some examples presented in the paper. The presented results show that the best equivocation codes produce a higher level of secrecy for the syndrome coding scheme than almost all best known error correcting codes. Interestingly, it is unveiled that some outstanding best known error correcting codes are also best equivocation codes.

I. Introduction

The wiretap channel, which was introduced by Wyner [1], is a physical layer model that captures the fundamentals of communication security. In this model, a transmitter, Alice, wishes to send confidential information to a legitimate receiver, Bob, in the presence of an eavesdropper, Eve. Wyner proposed a syndrome coding scheme to guarantee the security of the communication for a specific wiretap channel, where the main channel is error-free and the eavesdropper channel is a binary symmetric channel [2]. Wyner showed that as long as the size of the syndrome space is chosen to be smaller than the Shannon entropy of the binary symmetric channel, there exist codes that leak a vanishing proportion of information to the eavesdropper as the code length approaches infinity. In [3], Bennett strengthens Wyner's result in the sense that the channels are more general and the estimate of the number of information bits leaked to Eve is stronger. Cohen and Zemor [4] also provide a more refined analysis of the information leakage of syndrome coding for the wiretap channel.

The syndrome coding scheme is an important physical layer scheme to guarantee the secure communication that has been widely studied for various applications. For example, Cohen and Zemor [5] studied a generalization of Wyner's syndrome coding scheme which is applicable to noisy main and eavesdropper channels, for biometric applications. Reddy *et al.* [6] studied low-density parity-check (LDPC) codes for syndrome coding with applications to video coding systems. Suresh *et al.* [7] showed that duals of certain LDPC codes, when used in a syndrome coding scheme, provide strong secrecy over the binary erasure wiretap channel. Salim and Emina showed that network security can be achieved by using syndrome coding as an additional layer to a network code [8].

For the syndrome coding scheme, it is important to construct the good codes to guarantee the security of the communication. Numerous contributions have been made in designing good code

families such as BCH and Goppa codes [9] or by modifying good codes by concatenation [10], extension [11]–[14] or shortening [9]. Traditionally, good codes are designed for an error correcting system to combat noisy transmission channels and achieve reliable communication for different systems [15], [16]. A central property of an error correcting code is the minimum Hamming distance, d, which determines the number of independent errors that can be corrected and the resulting reliability of communication. The main objective in error correcting code design is to optimise one of the parameters n, k and d for given values of the other two [14]. The covering radius R_c is usually not considered a design parameter in traditional error correcting code design, but it is an indicator of codeword packing [9]. Tables of optimum binary codes are published in an online database by Grassl as Best Known Codes (BKC) in the form of tables of lower and upper bounds to d [17] for different code rates and code lengths up to 256 bits. With so much research attention historically, improvements to these tables are relatively rare.

Codes designed for error correction can also be used in syndrome coding, but it is shown below that the best performance is not generally achieved and better codes exist. The information rate of a syndrome coding scheme using an (n, k, d) linear code is $\frac{n-k}{n}$ and all possible binary vectors of length n may be transmitted. This is quite different from error correction coding applications where only codewords are transmitted and the information rate is k/n. For the syndrome coding, performance is measured in terms of the level of information theoretic secrecy, the equivocation rate, R_e [2]. The best codes have the highest value of equivocation rate for a given code length and code rate. It is convenient to use the parameter $m = n-k$, to represent the number of parity bits of the code. A best code for syndrome coding is represented by the functions: $n(m, R_e)$ - the shortest code for a given R_e and m and $R_e(n, m)$ - the highest equivocation rate for which a binary $(n, k = n - m)$ code exists. In contrast, a best code for error correction is represented by the function $n(m, d, R_c)$ - the longest code for a given d, m and R_c. Based on a list of parity check matrices of best performing codes for the syndrome coding scheme for given m and n, a Best Equivocation Codes (BEC) online database has been generated in the form of n, m and $R_e(n, m)$ [18].

In this paper we restrict the choice of codes to binary, linear codes. We present a numerical design technique to produce the best codes for the syndrome coding scheme as measured by the equivocation rate and compare these codes to the best known error correcting codes. The technique is limited practically to codes having around 35 parity bits, but this is not a major

limitation as near 100% secrecy is attainable with code lengths less than 256 bits [18]. The BKC error correcting code tables list codes less than 256 bits long, which make the comparisons possible. We also present an efficient recursive method to analyse the secrecy performance of any linear binary code, which is used in syndrome coding for the Binary Symmetric Channel (BSC). Finally, we present performance results of some best equivocation codes for the syndrome coding scheme and compare these to the best known codes. In almost all cases there is a substantial improvement.

II. Preliminaries: syndrome coding and the wiretap channel

The wiretap channel used in the syndrome coding scheme is shown as Figure 1, in which the main channel is noiseless and the eavesdropper channel is a binary symmetric channel with the crossover probability of α. This represents an abstraction of a typical bugging situation where the legitimate transmitter, Alice, and the legitimate receiver, Bob, employs a robust channel, and the eavesdropper, Eve, is listening, using a wireless bugging device with limited transmission power and antenna gain, typically with non line of sight radio propagation conditions so that she receives the transmissions with errors due to the poor communication channel. The secrecy capacity of this wiretap channel [1] is:

$$C_s = -\alpha \cdot \log_2 \alpha - (1-\alpha) \cdot \log_2 (1-\alpha) \tag{1}$$

which is the highest transmission rate that can be achieved whilst keeping communications secure from eavesdropping for this channel.

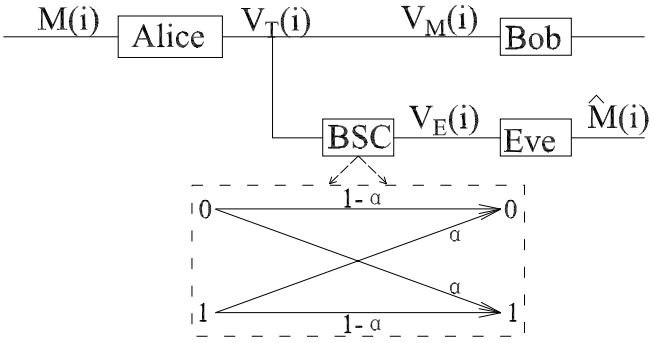

Figure 1. The wiretap channel model

Alice wishes to convey a sequence of independent and uniformly distributed m-bit binary messages, ***M(1)***$, \ldots,$ ***M(p)***, in which p is the block length of the messages, to the legitimate receiver. This sequence of m-bit messages, ***M(1)***$, \ldots,$ ***M(p)***, is encoded into a sequence of n-bit codewords, $\boldsymbol{V}_T(1), \ldots, \boldsymbol{V}_T(p)$. Bob observes the sequence of n-bit words, $\boldsymbol{V}_M(1), \ldots, \boldsymbol{V}_M(p)$, where

$$\boldsymbol{V}_M(i) = \boldsymbol{V}_T(i), \qquad i = 1, \ldots, p \tag{2}$$

and Eve observes the sequence of n-bit words, $\boldsymbol{V}_E(1), \ldots, \boldsymbol{V}_E(p)$, where

$$\boldsymbol{V}_E(i) = \boldsymbol{V}_T(i) + \boldsymbol{e}(i), \qquad i = 1, \ldots, p \tag{3}$$

and $\boldsymbol{e}(i)$ represents a n-bit error vector arising from the binary symmetric channel. Bob produces the sequence of original messages, $\boldsymbol{M}(1), \ldots, \boldsymbol{M}(p)$, from the main channel output sequence, $\boldsymbol{V}_M(1), \ldots, \boldsymbol{V}_M(p)$, whereas Eve produces a sequence of estimated messages, $\hat{M}(1), \ldots, \hat{M}(p)$, from the eavesdropper channel output sequence, $\boldsymbol{V}_E(1), \ldots, \boldsymbol{V}_E(p)$.

Traditionally, the syndrome coding scheme uses a $(n, k, 2t+1)$ linear block code, which is capable of correcting t errors, defined either by a $k \times n$ generator matrix $\boldsymbol{G}$ or by a $m \times n$ parity check matrix $\boldsymbol{H}$ [20]. It is a property of any linear block code that there exist 2^m distinct n-bit minimum weight error patterns, in which each pattern e_j produces a distinct syndrome S_j of the total 2^m syndromes, based on $S_j = e_j \times \boldsymbol{H}^T$, where $0 \leq j \leq 2^m - 1$. These minimum weight error patterns, e_j, are the coset leaders of the code and may be represented in a table of 2^m syndromes,S_j, and associated minimum weight error patterns, e_j. For perfect codes, the error patterns, e_j, in the syndrome table are all of the n-bit binary error vectors with a weight $w \leq t$. For non-perfect codes, some of the error patterns in the table are n-bit binary vectors with a weight $w \leq t$ and the remaining error patterns are n-bit binary vectors with a weight $w > t$. Regardless of whether or not the code is perfect, in syndrome coding all 2^m syndromes, S_j, are used to send messages. In the traditional syndrome coding, it is necessary to store an error pattern-syndrome look up table, which can be accessed by Alice, Bob and Eve. However, it is shown below that such a look up table is unnecessary and the parity check matrix of the code is sufficient.

With some elementary row and column operations ,every parity check matrix can be represented in a reduced echelon form [9]. In the following it is assumed that the parity check matrix

of the code is in the reduced echelon form with an identity sub-matrix of m parity bits followed by k information bit columns.

The encoding and decoding processes are as follows.

A. Encoder

Alice equates the m-bit messages to m-bit syndromes $S_T(i) = M(i)$ and produces n-bit vectors $V_T(i)$ from each m-bit message $M(i)$ at time i such that $V_T(i) \times \boldsymbol{H}^T = M(i)$ in three steps:

- Alice generates a random n-bit codeword $C_T(i)$ from a random, uniformly distributed k-bit vector $D_R(i)$:

$$C_T(i) = D_R(i) \times \boldsymbol{G} \tag{4}$$

- Alice forms an n-bit zero padded message vector,$V(i)$, consisting of the m-bit message $M(i)$ followed by k 0's.
- Alice then generates the transmitted n-bit vector $V_T(i)$ by adding the n-bit codeword $C_T(i)$ to the n-bit zero padded message vector

$$V_T(i) = C_T(i) + V(i) \tag{5}$$

In this way Alice produces n-bit vectors with the property that $V_T(i) \times \boldsymbol{H}^T = M(i)$.

Proof: The syndrome of $V_T(i)$, $S_T(i)$ satisfies

$$\begin{aligned} S_T(i) &= V_T(i) \times \boldsymbol{H}^T \\ &= C_T(i) \times \boldsymbol{H}^T + V(i) \times \boldsymbol{H}^T \\ &= 0 + M(i) \end{aligned} \tag{6}$$

$V(i) \times \boldsymbol{H}^T$ produces $M(i)$, which is due to the structure of $V(i)$ and $\boldsymbol{H}$. ■

The information rate of the syndrome coding scheme is $R = \frac{m}{n}$.

B. Legitimate decoder

Bob calculates the syndromes of the received vectors, $V_M(i)$ to determine the original messages:

$$S_{Bob}(i) = V_M(i) \times \boldsymbol{H}^T \tag{7}$$

As there are no transmission errors in Bob's channel $V_M(i) = V_T(i)$ and

$$S_{Bob}(i) = \boldsymbol{M}(i) \tag{8}$$

C. Eavesdropper decoder

Eve computes the syndromes associated with the vectors, $V_E(i)$, as follows:

$$S_{Eve}(i) = V_E(i) \times \boldsymbol{H}^T = V_T(i) \times \boldsymbol{H}^T + e(i) \times \boldsymbol{H}^T = \boldsymbol{M}(i) + S_e(i) \tag{9}$$

Eve's estimate of the message is given by:

$$\hat{M}(i) = S_{Eve}(i) = M(i) + S_e(i) \tag{10}$$

III. Calculation of the equivocation rate of a syndrome coding scheme

In syndrome coding scheme, the secrecy is measured by the eavesdropper decoder output equivocation, which can be calculated as follows:

$$\begin{aligned}
H(M(i)|\hat{M}(i)) &= H(M(i), \hat{M}(i)) - H(\hat{M}(i)) \\
&= H(M(i)) - H(\hat{M}(i)) + H(\hat{M}(i)|M(i)) \\
&= H(M(i)) - H(M(i) + S_e(i)) + H(M(i) + S_e(i)|M(i)) \\
&= m - m + 0 + H(S_e(i)|M(i)) \qquad (11) \\
&= H(S_e(i)) \qquad (12) \\
&= -\sum_{j=0}^{2^m-1} p(S_j)\log_2 p(S_j) \qquad (13)
\end{aligned}$$

in which, S_j denotes one of the total 2^m syndromes. The simplifications in equations (11) and (12) are due to $M(i)$ being uniformly distributed and independent of $S_e(i)$. The equivocation is calculated after deriving the probability mass function of the syndromes due to errors from the BSC, $p(S_j)$ and is a function of the parity check matrix of the code.

A. Packed integer representation of the parity check matrix

Historically many code design methods for linear codes, which have been presented in the literature, are based on construction of the parity check matrix, $\boldsymbol{H}$ [11]–[13]. In the following, the best equivocation codes are also constructed from the parity check matrix of the code.

Any column of a parity check matrix of a $(n, k, 2t+1)$ code, $\boldsymbol{H}$, can be represented by an integer, b_i, in the range 0 to $2^{n-k}-1$, in which i is the index of the column and $0 \leq i < n$. The parity check matrix of a $(n, k, 2t+1)$ code is defined by n integers, referred to as packed integers, and these integers can be in any order since the corresponding codes will all be equivalent [9]. Any code design of length n may be represented by these n packed integers. Since any parity check matrix can be put into reduced echelon form by elementary row and column matrix operations, the first m packed integers are fixed and the code design reduces to the determination of $n-m$ packed integers. Let us start with a general binary, reduced echelon parity check matrix:

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & \cdots & 0 & a_{m0} & \cdots & a_{(n-1)0} \\ 0 & 1 & \cdots & 0 & a_{m1} & \cdots & a_{(n-1)1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & a_{ij} & \vdots \\ 0 & 0 & \cdots & 1 & a_{m(m-1)} & \cdots & a_{(n-1)(m-1)} \end{pmatrix} \tag{14}$$

in which $a_{i,j}$ takes a value of 0 or 1 ,i and j denote the index of the rows and the columns respectively, $0 \leq j \leq m-1$ and $m \leq i \leq n-1$. Each column may be represented as a packed integer defined as $b_i = \sum_{j=0}^{m-1} a_{ij} \cdot 2^j$. The systematic packed integer form of the parity check matrix is $[1, 2, \cdots, 2^{m-1}, b_m, \cdots, b_{n-1}]$. For example, the parity check matrix of the (15,7,5) BCH code is:

$$\boldsymbol{H} = \begin{pmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,0\,0\,1\,1 \\ 0\,1\,0\,0\,0\,0\,0\,0\,1\,1\,0\,1\,1\,1\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0\,0\,1\,1\,0\,1\,1\,1 \\ 0\,0\,0\,1\,0\,0\,0\,0\,1\,1\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,1\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,1\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,1\,0\,1 \end{pmatrix} \tag{15}$$

which in packed integer representation is:

$$\boldsymbol{H} = [1, 2, 4, 8, 16, 32, 64, 128, 27, 47, 117, 202, 166, 71, 149] \tag{16}$$

B. Analysis on the syndrome probability distribution

Equation (13) shows that it is necessary to get the probability mass function of the syndromes, $p(S_j)$, for the calculation of the equivocation in BSC. There are two ways to evaluate $p(S_j)$, in

which $0 \leq j \leq 2^m - 1$.

1) Traditional evaluation method of the syndrome probability distribution: There are 2^n possible ways in which error patterns, $e(i)$, occur in each transmitted n-bit vector. These error patterns occur with probability:

$$p(e(i)) = \alpha^{w(i)} \cdot (1-\alpha)^{n-w(i)} \tag{17}$$

where $w(i)$ is the weight of $e(i)$. Each error pattern results in one of the 2^m syndromes, which is produced as follows:

$$S_e(i) = e(i) \times \boldsymbol{H}^T \tag{18}$$

Any error pattern, which is a codeword, has a syndrome equal to zero. As the code is linear, for each particular syndrome, S_j, there are 2^k different error patterns, $e(i)$, that produce the same syndrome and the probability of each syndrome due to all possible such error patterns is given by

$$p(S_j) = \sum_{i=0}^{2^n-1} p(e(i))\delta(S_e(i) - S_j) \tag{19}$$

where S_j denotes a specific syndrome vector in all of the 2^m syndromes, $0 \leq j \leq 2^m - 1$, and $\delta()$ is the Dirac function. We have

$$H(S_e(i)) = -\sum_{j=0}^{2^m-1} p(S_j) \log_2 p(S_j) \tag{20}$$

This method for evaluating the equivocation is computationally manageable for short codes ($n < 40$), but for the long codes it is not practical because it involves the evaluation of 2^n error patterns, an exponential function of the code length. However the probability distribution of the syndromes may be determined recursively and this also provides some insight into code construction by code extension.

2) Recursive evaluation of the syndrome probability distribution: Let us represent the parity check matrix of the code in the packed integer format. The packed integers directly correspond to syndrome values in that a single bit error in a transmitted codeword results in a syndrome equal to the packed integer of the column corresponding to the bit error position. As the codes are linear, any combination of bit errors produces a syndrome equal to the modulo 2 sum of the packed integers corresponding to each column of the parity check matrix in which a bit error occurs.

The following theorem shows that the probability mass function of the syndrome, $p(S_j)$, is a function of the parity check matrix of the code and the crossover probability of the BSC.

Theorem 1: The probability mass function (pmf) of S_j for $j = 0$ to $2^m - 1$ may be defined as $p(S_j) = \beta(j)$ where $\beta(j)$ are coefficients of the probability generating function using the Z transform, $p_z(\mathcal{S})$ and $p_z(\mathcal{S})$ only depends on the columns of the parity check matrix and α.

$$p_z(\mathcal{S}) = \sum_{j=0}^{2^m-1} \beta(j) Z^j = \prod_{i=0}^{n-1} \left((1-\alpha) + \alpha Z^{b_i}\right) \tag{21}$$

where b_i are the packed integer representations of the columns of the parity check matrix and exponent sums of powers of Z are modulo 2 sums.

Proof: Any error pattern may be represented as a sum of single bit error events:

$$\begin{aligned} e(i) =& [e^1 \ e^2 \ \cdots \ e^n] \\ =& [e^1 \ 0 \ \cdots \ 0] + [0 \ e^2 \ \cdots \ 0] + \cdots + [0 \ 0 \ \cdots \ e^n] \end{aligned} \tag{22}$$

in which each term denotes a single bit error event where $e^i = 1$ with probability α and $e^i = 0$ with probability $1 - \alpha$.

The linearity of the syndrome coding scheme means that the syndrome resulting from any error pattern is the sum of the syndromes for each bit error position:

$$S_e(i) = e(i) \times \boldsymbol{H}^T = [e^1 \ e^2 \ \cdots \ e^n] \times \boldsymbol{H}^T \tag{23}$$

$$= b_1\delta(e^1 - 1) \oplus b_2\delta(e^2 - 1) \cdots \oplus b_n\delta(e^n - 1) \tag{24}$$

in which $\oplus$ denotes the modulo 2 sum. Since the probabilities of e^1, e^2, $\cdots$,e^n are independent, the probability of $S_e(i)$ is the product of the probabilities of n separate error events. Adding together the coefficients of same powers of Z results in the coefficients, β_j and reduces the number of terms from 2^n to 2^m. ■

Theorem 1 gives us some insight into code construction by code extension.

C. Code construction by code extension

There is a well known relationship between the parity check matrix and the minimum Hamming distance of the code from first studies of error correction coding [19]:

Property 1: A linear code has d minimum distance if and only if its parity check matrix has d linearly dependent columns but no set of $d - 1$ or fewer, linearly dependent columns.

Based on *Property*1, any packed integer of the parity check matrix cannot be equal to the modulo 2 sum of any combination of $d-2$ packed integers or fewer integers of the parity check matrix. This is apparent by observing that the syndrome of a codeword is equal to zero. If any packed integer of the parity check matrix were equal to the modulo 2 sum of any combination of $d-2$ packed integers or fewer integers then the resulting combination would have a syndrome equal to zero and therefore be a codeword of weight less than d, an obvious contradiction.

A corollary of *Property*1 is that every error pattern of weight less than $t+1 = \lfloor \frac{d+1}{2} \rfloor$ has a distinct syndrome. For syndrome coding this means that the syndrome probability mass function considering only error events with weight less than $t+1$ will be maximally flat. However as higher weight error events all the way up to weight n can occur, other properties of the code other than d are involved.

The code construction method for obtaining codes with good equivocation is based on the observation that the syndrome probability mass function of a code extended in length is a function of the probability mass function of the original code and good equivocation codes produce good extended codes and this is elaborated below. BKC's have also been obtained by code extension. Bouyukliev's code construction method extends a code by excluding all linear combinations of up to $d-2$ columns of the parity check matrix [14] exploiting *Property*1. The inverting construction Y_1, proposed by Edel adds an additional parity bit and extends a (n,k,d) code to a $(n+x, k+x-1, d)$ code, by extending the parity check matrix [11] similar to Bouyukliev.

Theorem 1 gives us some insight into code construction by code extension. If the columns of the parity check matrix of the shortened code with a length r are taken consecutively from $i=0$ to $r-1$, in which i is the index of the column, and the Z transform the probability generating function of the shortened code is

$$p_z(\mathcal{S}_r) = \prod_{i=0}^{r-1} \left((1-\alpha) + \alpha Z^{b_i} \right) \tag{25}$$

then the Z transform of the probability generating function the code of length $r+1$ is

$$p_z(\mathcal{S}_{r+1}) = \prod_{i=0}^{r} \left((1-\alpha) + \alpha Z^{b_i} \right) = p_z(\mathcal{S}_r)((1-\alpha) + \alpha Z^{b_r}) \tag{26}$$

By denoting the coefficients of the shortened code of length r, as $\beta_r(i)$ then

$$p_z(\mathcal{S}_r) = \sum_{j=0}^{2^m-1} \beta_r(j) Z^j \tag{27}$$

and

$$p_z(\mathcal{S}_{r+1}) = (1-\alpha)\sum_{j=0}^{2^m-1}\beta_r(j)Z^j + \alpha\sum_{j=0}^{2^m-1}\beta_r(j)Z^{j\oplus b_i} \tag{28}$$

which simplifies to

$$p_z(\mathcal{S}_{r+1}) = (1-\alpha)p(\mathcal{S}_r) + \alpha\sum_{j=0}^{2^m-1}\beta_r(j)Z^{j\oplus b_i} \tag{29}$$

It is apparent that the syndrome probability generating function in Z transform of the code of length $r+1$ is equal to the syndrome probability generating function in Z transform of the shortened code of length r, weighted by $1-\alpha$ plus a permuted syndrome probability generating function in Z transform of the shortened code of length r, weighted by α. The permutation arises from the results of the modulo 2 additions $j \oplus b_i$. In general, the more uniform the syndrome pmf of a code is, the more uniform the syndrome pmf of the extended code will be.

This leads to the conclusion that the syndrome pmf of the code may be obtained recursively, starting with the generating function $p_z(\mathcal{S}_1)$ determining $p_z(\mathcal{S}_2)$ then $p_z(\mathcal{S}_3)$ through to $p_z(\mathcal{S}_n)$. It is also apparent that good equivocation codes will also produce good equivocation codes when extended in length.

IV. Code design for syndrome coding

We are now in a position to consider code design for a syndrome coding scheme for the wiretap channel. To construct the best codes for the syndrome coding scheme, we want the equivocation of the eavesdropper to be as high as possible, i.e. the pmf of the syndromes should be as uniform as possible. Since the eavesdropper channel is a binary symmetric channel, for relatively low values of the error probability, α, the equivocation is dominated by error events of low weight. We note the following design considerations:

1) For the BSC, low weight error events have higher probabilities, so they dominate the pmf of the syndrome. If the low weight error events produce distinct syndromes, this helps the pmf of the syndromes to become more uniform. This has a connection to the minimum Hamming distance of the code, d, because error events of weight up to $\frac{d-1}{2}$ all have distinct syndrome sums. Best known codes, which have the highest d, are likely to give good performance in the syndrome coding scheme, but are not necessarily the best codes,

because all error patterns, regardless of weight, contribute towards the syndrome pmf of the code.

2) With the parity check matrix in reduced echelon form, that is in systematic packed integer form, the packed integers of any of the information bits cannot have a weight less than $d-1$; otherwise the codeword formed from that information bit alone will have weight less than d, where d is the minimum Hamming distance of the code.
3) There is at least one syndrome that is produced by a weight R_c error event and no lower weight error event, where R_c is the covering radius of the code. A good equivocation code will tend to have low covering radius compared to the minimum Hamming distance, d. Perfect codes have low covering radius and are good equivocation codes.
4) A shortened best known code has a bounded covering radius $R_c \geq (d-1)$, because there exists at least one coset leader of the shortened code of weight $d-1$ or greater; otherwise the extended code formed by adding this coset leader to the generator matrix of the code will produce a code with minimum Hamming distance less than d, a contradiction. It is unlikely that shortened best known codes will be best equivocation codes.
5) If columns of the parity matrix are repeated, a weight 2 error event will produce a zero syndrome, which makes the pmf of the syndrome non-uniform. To get a good equivocation code, the parity check matrix should have no repeated integers, unless the code is very long compared to the number of parity bits.

Following these observations, a code design algorithm has been formulated, that produces best equivocation codes as follows:

INPUT: C_{in}-Set of l inequivalent [1] highest equivocation rate (n, m) codes with parity check matrices in systematic packed integer form $\boldsymbol{H}$.

OUTPUT: C_{out}- Set of all extended inequivalent $(n+1, m)$ codes, which are ranked by equivocation in descending order.

For each code C in C_{in}, the extension steps are as follows:

1) Preset a status array $a[i] = 1$, where $1 \leq i \leq 2^{n-k}$.
2) If i is equal to any integer in $\boldsymbol{H}$, set $a[i] = 0$.

[1]If the codes have different equivocation rates, they are inequivalent codes.

3) For all j such that $a[j] = 1$, extend $\boldsymbol{H}$ with one integer, equal to j, to $\boldsymbol{H} = [\boldsymbol{H}, j]$. Eliminate all equivalent codes and evaluate the equivocation[2] for each remaining code.
4) Rank the inequivalent codes by their equivocation in descending order, C_{out}, and select the first l codes. This is termed the best codes subset. These are used as the INPUT for the next extension round.

For a given m, all possible $(m+1, 1)$ codes are used as the initial INPUT with the parity check matrices: $[1, 2, \cdots, 2^{m-1}, i]$, for all i, $1 \leq i \leq 2^{n-k}$. For each round as n is incremented, the best equivocation codes are obtained, provided that the best codes subset is sufficiently large. The syndrome probability generating function in Z transform of each (n, m) code, $p_z(\mathcal{S})_{in}$, is stored and the syndrome probability generating function in Z transform for each extended $(n + 1, m)$ code, $p_z(\mathcal{S})_{out}$, is determined using equation (29) which makes for a fast algorithm.

Usually, the best extended code will be derived from the highest ranked code from the previous round but this is not always the case. The size of the best codes subset, l, needs to be large enough that no better extended code is missed. For short codes it is possible to store all inequivalent codes in the best codes subset and it is proven that the highest ranked code is the best (n, m) equivocation code.

To estimate the size required of the best codes subset, for a given number of parity bits, l is assigned (e.g. $l = 20000$) and the code is extended in length by several bits. For the highest ranked code of length n, the ranked position p_{n-1} is determined of its parent code of length $n - 1$ from the previous round. Statistics of p_{n-1} are collected for each code length and the subset size l is maintained to be greater by a significant margin to the maximum value of p_{n-1}, i.e. $l \geq \max p_{n-1}$. For $m = 6$ and $m = 7$, for all code lengths of $n \leq 70$, the maximum p_{n-1} values are 25 and 275 respectively. It is observed that the greater the length of the code, the higher the possibility of $p_{n-1} = 1$, i.e. the extended best equivocation code of length n produces the best equivocation code of length $n + 1$.

V. Results

Following the procedure above, the best equivocation codes have been determined for a given number of parity bits and code length. There are too many codes to present, so we only present

[2]The equivocation is evaluated for the binary symmetric eavesdropper channel for a given error probability.

here some representative codes. The codes are listed in an on-line database [18] in packed integer format. The minimum Hamming distance, covering radius and equivocation rate for a BSC error probability of $\alpha = 0.05$ are given for each code.

As examples, Table I lists the highest rate codes, in packed integer format, which provide at least 90% secrecy when used in syndrome coding for a BSC error probability of $\alpha = 0.05$. Interestingly, in general nearly all of the best equivocation codes do not coincide with the best known codes. However these best equivocation codes do have a respectable minimum Hamming distance and covering radius.

With reference to the on-line database [18], there are a few best known codes that are also best equivocation codes, notably all of the perfect codes. The perfect codes are the repetition codes, the Hamming codes and the Golay code. The reason for this coincidence is that these codes have the highest minimum Hamming distance, d, and are optimally packed codes with the lowest possible covering radius. It is notable that there is a small number of other best known codes that are also best equivocation codes and we term these best known codes as well packed codes. A feature of these well packed codes is that for a given number of parity bits there is no known longer code with the same minimum Hamming distance, d. Table II gives the code parameters of the well packed BKC for various values of m. We have observed that some of the shorter well packed BKC coincide with the best equivocation codes for given m and given k.

It is also interesting to note that each well packed BKC of length n is followed by a poorly packed BKC of length $n + 1$. These BKC's are termed the worst packed BKC's. It is found that the corresponding BEC's have much higher equivocation rates than that of the worst packed BKC's.

Figure 2 compares the equivocation rate of best known codes and best equivocation codes for increasing code rate, for codes with 26 parity bits at different values of α. It shows that the BECs have higher equivocation than BKCs not only for $\alpha = 0.05$ but also for other values of α. The dashed line gives the secrecy capacity of the channel. It is readily apparent that the BEC's outperform the BKC's particularly in the lower rate region.

As the number of parity bits increase, codes get longer for the same information rate and the equivocation moves closer to the secrecy capacity limit, C_s, which is applicable to infinite length codes. This is shown in Figure 3 for $m = 6$ through to $m = 26$.

Table I

BEST EQUIVOCATION CODES THAT ACHIEVE 90% SECRECY IN SYNDROME CODING FOR $\alpha = .05$

m	n	d	Rc	Re	packed integer parity check matrix
4	25	2	1	0.906251	1 2 4 8 1 2 3 3 4 5 5 6 6 7 8 9 10 11 11 12 12 13 14 15 15
5	28	3	2	0.907646	1 2 4 8 16 3 5 6 7 9 11 12 13 14 15 17 18 19 20 22 23 24 25 26 27 28 29 30
6	31	4	3	0.908721	1 2 4 8 16 32 7 11 13 14 19 21 22 25 26 28 31 35 37 38 41 42 44 47 49 50 52 56 59 61 62
7	35	3	3	0.907273	1 2 4 8 16 32 64 7 14 25 26 28 31 38 43 51 53 56 62 70 75 83 85 88 93 100 103 104 110 112 118 121 122 124 127
8	38	4	3	0.904458	1 2 4 8 16 32 64 128 26 35 46 61 79 83 86 89 98 101 105 126 131 143 156 166 171 183 185 193 198 200 203 205 223 231 236 243 244 250
9	41	4	3	0.900208	1 2 4 8 16 32 64 128 256 31 45 60 100 113 122 143 153 156 181 191 194 204 214 218 281 291 302 331 332 344 370 383 394 397 403 418 427 440 478 485 488
10	45	4	3	0.904383	1 2 4 8 16 32 64 128 256 512 29 87 109 191 203 221 229 301 344 350 379 409 426 435 467 468 495 501 527 539 596 613 637 661 675 716 755 787 820 858 879 897 969 994 1008
11	48	4	3	0.902358	1 2 4 8 16 32 64 128 256 512 1024 82 95 103 137 170 340 376 415 416 456 486 579 647 734 754 811 829 887 964 977 1079 1082 1227 1239 1302 1370 1420 1505 1546 1596 1710 1828 1863 1929 1938 1975 2041
12	51	4	3	0.900428	1 2 4 8 16 32 64 128 256 512 1024 2048 108 214 225 327 415 460 777 814 862 939 978 998 1059 1116 1206 1450 1606 1663 1677 1845 2243 2355 2456 2554 2951 2964 2985 3119 3414 3467 3582 3731 3802 3812 3819 3907 3928 3950 4049
13	55	6	5	0.905225	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 309 345 618 687 690 981 1207 1236 1374 1380 1653 1897 1962 2373 2414 2472 2639 2748 2760 3306 3794 3924 4315 4735 4746 4807 4828 4887 4944 5278 5496 5520 6225 6612 7263 7273 7411 7449 7588 7848 8007 8137
14	58	5	4	0.902823	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 463 659 1306 1903 2017 2420 2623 2681 2774 3009 3018 3372 3462 3727 3848 4807 4978 5102 5286 5495 5755 6833 7557 7706 8435 9814 10158 10732 10769 10991 11355 11837 12664 12788 13759 13762 13880 15043 15411 15498 15640 15910 16068 16352
16	64	5	4	0.9	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 2026 2174 6623 9685 10416 10555 11116 11896 12221 12999 14403 14586 14926 19995 20228 20788 23508 23688 23953 25340 27319 27881 28822 30545 31261 31842 32143 33695 35079 35300 40115 41850 42821 43146 46587 46930 48617 48950 50224 50393 52924 53375 53724 55136 57952 61511 62572 65419
18	71	6	5	0.903035	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 11955 13936 32282 32999 38800 39250 42924 47840 55393 56951 58478 60921 69584 75429 81270 92272 93719 98994 101001 102495 107633 112602 118747 135116 148158 154496 159264 161706 168209 175488 176485 180456 187783 193685 194276 195402 203311 206292 207540 207680 208910 210560 225118 230979 235516 238390 247552 252600 254562 254936 256876 258694 260096
20	77	8	7	0.900739	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 262144 524288 41772 53076 70933 85654 95837 102185 121984 143743 159055 160966 162356 177792 186954 191057 197822 208965 259410 268572 272682 296292 308553 313067 320896 366336 371328 405616 442424 445952 450779 457021 475870 491392 532261 539043 560237 602893 616832 625083 632370 646431 661559 680024 692328 692902 701447 716691 741146 746583 749617 756858 782531 799616 814830 923790 985856 996729 1044736
22	84	8	7	0.903293	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 262144 524288 1048576 2097152 14191 111264 120844 135402 156066 193837 297854 309690 330532 338475 430909 567303 682894 714276 780924 869995 905402 923298 982920 1351071 1376188 1380394 1419610 1487155 1551602 1884223 1964241 2073026 2141889 2289013 2314096 2350121 2376601 2501526 2638285 2645293 2687586 2703451 2758616 2779212 2825279 2893319 3047974 3082337 3148814 3167196 3172106 3212496 3244862 3367783 3378339 3595701 3671892 3675563 3727842 3733852 3753502 3954669 3966529 3984378 3994343 4013090
24	90	7	6	0.901784	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 262144 524288 1048576 2097152 4194304 8388608 573185 576055 605300 649314 701023 902444 935950 972344 1107341 1216620 1391649 1700311 1849660 2435061 2801845 3274804 3727802 3857484 4410637 4447108 4642815 4754320 4865882 4933529 4981695 5087490 5170378 5176601 5211409 5493873 6813668 6982303 7139876 7894909 7933395 8894748 9354907 9466165 10033669 10290550 10320865 10887849 11095914 11204875 11347996 11485683 11727110 11796941 12233544 12889002 13314240 13707676 14157669 14737865 14795911 14837286 14923311 15030096 15150513 15341837 15855350 15923713 15972347 16494372 16540142 16554318
26	97	8	6	0.904462	1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 262144 524288 1048576 2097152 4194304 8388608 16777216 33554432 1160185 3419307 4492287 9395871 10093907 11068401 11840460 12139956 14534737 14816701 17641732 17992524 18692090 20131180 20858618 22460065 22562674 23680920 23791323 24279912 25579531 26280684 26831716 27845706 28552180 28936460 32368082 32564036 34048695 34254884 35283464 35985048 35986053 36286430 38278580 39088097 40262360 41667867 43800122 44014004 44811375 46058364 46251188 47221140 47347900 47361840 47389686 47437732 47592948 47926580 48559824 49569716 52561368 53663432 54332396 54744803 54780620 55996779 56474461 56772244 57104360 57872920 58140380 58689469 60935573 61409218 62624620 63555923 64592814 65128072 65473972

Table II

WELL PACKED BKC

m	Well packed BKC, (n,k,d)
4	(5,1,5),(8,4,4),(15,11,3)
5	(6,1,6),(16,11,4),(31,26,3)
6	(7,1,7),(8,2,5),(32,26,4)
7	(8,1,8),(9,2,6),(11,4,5),(64,57,4)
8	(9,1,9),(10,2,6),(11,3,5),(128,120,4),(255,247,3)
9	(10,1,10),(11,2,7),(18,9,6),(23,14,5)
10	(11,1,11),(12,2,8),(15,5,7),(24,14,6),(33,23,5)
11	(12,1,12),(16,5,8),(23,12,7),(34,23,6),(47,36,5)
12	(13,1,13),(14,2,9),(24,12,8),(48,36,6),(65,53,5)
13	(14,1,14),(15,2,10),(25,12,8),(27,14,7),(66,53,6),(81,68,5)
14	(15,1,15),(16,2,10),(17,3,9),(28,14,8),(31,17,7),(82,68,6),(128,114,5)
15	(16,1,16),(17,2,11),(18,3,10),(20,5,9),(32,17,8)(37,22,7),(129,114,6),(151,136,5)
16	(17,1,17),(18,2,12),(21,5,10),(23,7,9),(38,22,8),(47,31,7),(152,136,6)
17	(18,1,18),(19,2,12),(20,3,11),(24,7,10),(27,10,9),(48,31,8),(63,46,7)
18	(19,1,19),(20,12,13),(21,3,12),(23,5,11),(28,10,10),(31,13,9),(64,46,8),(68,50,7)
19	(20,1,20),(21,2,14),(24,5,12),(26,7,11),(32,13,10),(35,16,9),(69,50,8),(88,69,7)
20	(21,1,21),(22,2,14),(27,7,12),(31,11,11),(36,16,10),(41,21,9),(89,69,8),(95,75,7)
21	(22,1,22),(23,2,15),(24,3,13),(32,11,12),(33,12,11),(42,21,10),(45,24,9),(96,75,8),(128,107,7)
22	(23,1,23),(24,2,16),(25,3,14),(27,5,13),(34,12,12),(36,14,11),(46,24,10),(49,27,9),(129,107,8),(155,133,7)
23	(24,1,24),(25,2,16),(28,5,14),(29,6,13),(37,14,12),(47,24,11),(50,27,10),(54,31,9),(156,133,8),(162,139,7)
24	(25,1,25),(26,2,17),(27,3,15),(30,6,14),(32,8,13),(48,24,12),(55,31,10),(64,40,9),(163,139,8)
25	(26,1,26),(27,2,18),(28,3,16),(31,6,15),(33,8,14),(34,9,13),(49,24,12),(65,40,10),(72,47,9)
26	(27,1,27),(28,2,18),(32,6,16),(35,9,14),(38,12,13),(50,24,12),(51,25,11),(73,47,10),(77,51,9)
27	(28,1,28),(29,2,19),(33,6,16),(35,8,15),(39,12,14),(40,13,13),(52,25,12),(63,36,11),(78,51,10),(94,67,9)
28	(29,1,29),(30,2,20),(31,3,17),(36,8,16),(37,9,15),(41,13,14),(44,16,13),(64,36,12),(69,41,11),(95,67,10),(128,100,9)
29	(30,1,30),(31,2,20),(32,3,18),(38,9,16),(40,11,15),(45,16,14),(46,17,13),(70,41,12),(71,42,11),(129,100,10),(135,106,9)
30	(31,1,31),(32,2,21),(33,3,18),(34,4,17),(41,11,16),(44,14,15),(47,17,14),(48,18,13),(72,42,12),(136,106,10),(142,112,9)

VI. CONCLUSIONS

We considered the code design problem for syndrome coding from the information theoretic security view and presented a code construction method for best equivocation codes. In this way, an online table of constructed BEC's in packed integer format has been produced. Code examples have been given of the highest rate BEC's that achieve 90% secrecy to an eavesdropper using

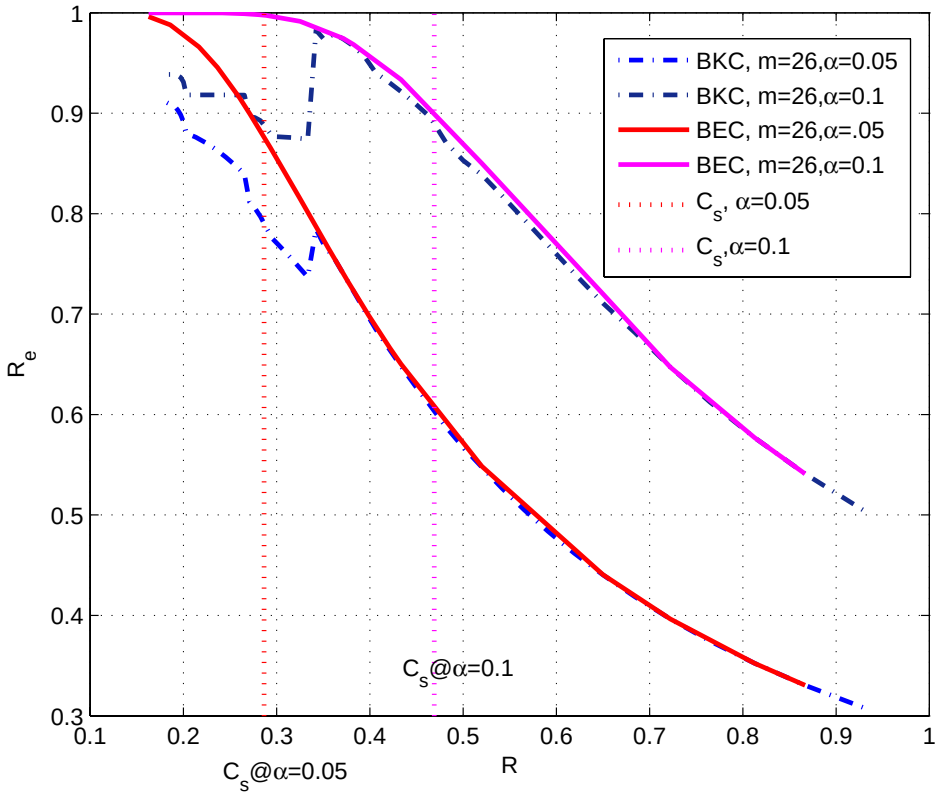

Figure 2. Comparison of the equivocation rate of best equivocation and best known codes having 26 parity bits

the BSC with an error probability of 0.05. Unlike the traditional case, the presented encoding method does not require a syndrome look up table. This encoding method also provides a new interpretation of syndrome coding in which the message is impressed on the parity bits of a random codeword. In effect the m parity bits of an n-bit random codeword are used to scramble an m-bit message necessitating the error free recovery of all n bits for an eavesdropper in order to determine the m bit message.

An efficient recurrent method was also presented for the calculation of the probability mass function of the syndromes of a code from the parity check matrix of the code, which in turn enables the secrecy achieved by the code, the equivocation rate, to be determined. It was shown that apart from the perfect codes, the best equivocation codes rarely coincide with the best known error correcting codes and that usually a best equivocation code will provide greater secrecy than

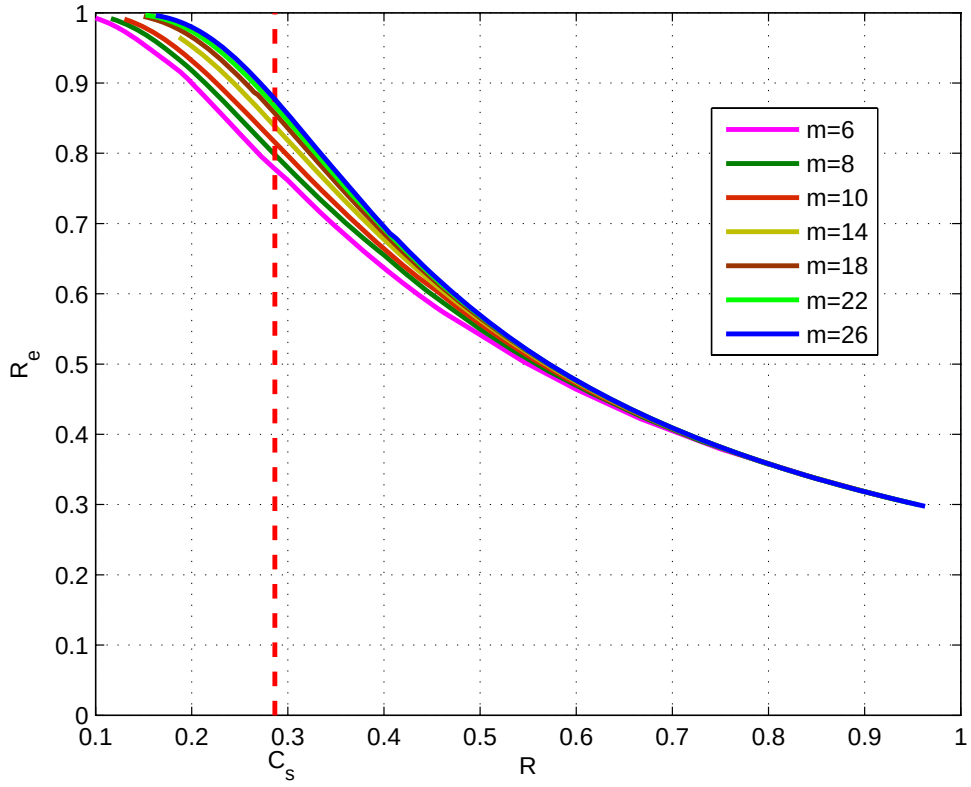

Figure 3. Equivocation rate vs. information rate for Best Equivocation Codes as a function of m

using a best known error correcting code with the same parameters.

REFERENCES

[1] Wyner, A.D.:'The wire-tap channel', Bell Syst. Tech. J., 1975, 54, (8), pp. 1355-1367

[2] Ozarow, L.H., Wyner, A. D.:'Wire-tap channel II', Bell Syst. Tech. J., 1984, 63, (10), pp. 2135-2157

[3] Bennett, C.H., Brassard, G., Crepeau, C., Maurer, U. M.:'Generalized privacy amplification', IEEE Transaction on Information Theory, 1995, 41, (6), pp. 1915-1923

[4] Cohen, G., Zemor, G.:'Syndrome-coding for the wiretap channel revisited', Proceedings of 2006 IEEE Information Theory Workshop(ITW'06), Chengdu, China, October 2006, pp. 33-36

[5] Cohen, G., Zemor, G.:'Generalized coset schemes for the wire-tap channel: application to biometrics', Proceedings. International Symposium on Information Theory 2004, Chicago, USA, June 2004, pp. 46

[6] Reddy, S., Aparna, P., David, S.:'Syndrome coding of video with LDPC codes', 9th International Conference on Signal Processing, 2008, Beijing, China, October 2008, pp. 1985-1988

[7] Suresh, A.T., Subramanian, A., Thangaraj, A., Bloch, M., McLaughlin, S.W.:‘Strong secrecy for erasure wiretap channels’, Proceedings of 2010 IEEE Information Theory Workshop (ITW’10), Dublin, Ireland, September 2010, pp. 1-5

[8] Salim, E.R.A., Emina, S.:‘On wiretap networks II’, IEEE International Symposium on Information Theory, 2007, Nice, France, June 2007, pp. 551-555

[9] MacWilliams, F.J., Sloane, N.J.A.: ‘The theory of error-correcting codes’,(New York: North-Holland publishing company, 3rd edn, 1981)

[10] Blokh, E. L., Zyablov, V.V.:‘Coding of generalized concatenated codes’, Probl. Inform. Transm., 1974, 10, (3), pp. 218-222

[11] Edel, Y.:‘Inverting construction Y1’, IEEE Transactions on Information Theory, September 1998, 44, (5), pp. 1993-1996

[12] Alltop, W.O.:‘A method for extending binary linear codes’,IEEE Transactions on Information Theory, November 1984, 30, (6), pp. 871-872

[13] Sloane, N.J.A., Reddy, S.M., Chen, C.L.: ‘New binary codes’, IEEE Transactions on Information Theory, 1972, 18, (4), pp.503-510

[14] Bouyukliev, I.G., Jacobsson, E.:‘Results on binary linear codes with minimum distance 8 and 10’, IEEE Transactions on Information Theory, 2011, 57, (9), pp. 6089-6093

[15] Sanaei, A., Ardakani, M.:‘LDPC code design considerations for non-uniform channels’, IEEE Transactions on Communications, January 2010, 58, (1), pp. 101-109

[16] Li, J., Yuan, J., Malaney, R., Azmi, M.H., Xiao, M.:‘Network coded LDPC code design for a multi-source relaying system’, IEEE Transactions on Wireless Communications, May 2011, 10, (5), pp. 1538-1551

[17] Grassl, M.:‘Bounds on the minimum distance of linear codes and quantum codes’, 2007, online, Available: http://www.codetables.de

[18] Zhang, K.:‘Best equivocation rate codes’, 2013, online, Available: http://www.it.pt/auto_temp_web_page_preview.asp?id=1219

[19] Peterson, W.W.:‘Error-correcting Codes(first edition)’, (MIT Press, 1961)

[20] Lin, S., Costello, D.J.:‘Error control coding (second edition)’, (Pearson Education, 2004, 2nd edn), pp. 100-231